Author's response to reviews

Title: Psychosis candidate genes in the prefrontal cortex: Meta-analysis of gene expression microarray studies

Authors:

Kwang H Choi (choik@stanleyresearch.org)
Michael Elashoff (mike@elashoffconsulting.com)
Brandon W Higg (brandon@elashoffconsulting.com)
Jonathan Song (sjonathan@comcast.net)
Sanghyeon Kim (kims@stanleyresearch.org)
Sarven Sabunciyan (ssabunc1@jhmi.edu)
Suad Diglisic (diglisics@stanleyresearch.org)
Robert H Yolken (rhyolken@aol.com)
Michael B Knable (knablem@stanleyresearch.org)
E Fuller Torrey (torreyf@stanleyresearch.org)
Maree J Webster (websterm@stanleyresearch.org)

Version: 2 Date: 12 September 2008

Author's response to reviews: see over
Sept. 12, 2008

Dear Editor and Reviewers,

Please find the revised version of MS: 1711616397203642 entitled: “Putative psychosis genes in the postmortem prefrontal cortex: Combined analysis of gene expression microarray studies” for your consideration. We thank the Reviewers for their detailed and insightful comments, and hope that this revised manuscript will be appropriate for acceptance in the BMC Psychiatry.

Response to reviewers

We are pleased that Reviewers noted the novelty and relevance of this study to the psychiatric genomics field, as well as the benefits of cross-study analysis. Reviewers also noted the complexity of the data analysis on multiple microarray studies, and with their encouragement we have improved our manuscript.

Below we provide each Reviewer’s comments in italics, and our response in normal text. When appropriate, we note how the revised manuscript has been changed.

Reviewer's report 1

Choi et al present an analysis across seven gene expression datasets obtained from the Stanley Genomics database, to identify expression changes between psychotic and non-psychotic patients. They identify several genes including metallothionein genes and some neuropeptides.

Major Compulsory revisions

1. In an earlier publication by the authors (Elashoff, et a., J Mol Neurosci. (2007)), a similar analysis is carried out on BPD and control data, using some of the same data stratified in a different way. In that study the authors discuss the potential impact of the samples having subjects in common, and the use of a degree of freedom correction factor to correct for study interdependencies (this correction is not described in any detail in the earlier paper). It is not clear why a similar correction was not deemed necessary in the current study. The authors should address this lack of independence head-on and thoroughly describe any correction they do decide to make. As it stands,
the pvalues they computed cannot be taken seriously and the results are of questionable statistical significance.

We apologize for the lack of clarity in the previous manuscript and we provided information to explain the correction factor analysis that was used to account for dependencies between study samples in our revised manuscript (page 7). We have used the same method in the previous paper (Elashoff et al. 2007). This permutation test allowed us to identify the appropriate degrees of freedom, and a correction factor was applied to adjust the over-estimated sample sizes in our analysis.

2. I'm concerned that the authors are dredging some of the same data they used in Elashoff et al. There they studied the bipolar disease effect. Now they are restratifying the same samples based on psychosis (and adding additional samples), which is obviously not independent of BPD.

We agree with the Reviewer that we are stratifying our subject population by a major psychiatric phenotype, psychosis. As we described in the introduction, psychosis is becoming an important psychiatric phenotype besides the traditional dichotomy between schizophrenia and bipolar disorder. This idea is discussed in recent papers showing increased interest on psychosis-related disorders (Owen et al. 2007, Pearlson and Folley 2008). We have comprehensive clinical information on each of these subjects and this allows us to investigate the data in different ways to extract gene expression patterns that can be used as putative markers for psychosis. Please note that most of schizophrenia datasets have not been analyzed and published before. Therefore, we believe that utilizing this large dataset to answer questions in a different way (psychosis) is justifiable.

In the previous paper (Elashoff et al. 2007), we used 14 different studies that included 3 additional non-Affymetrix microarray platforms. However, we are currently limiting our study to the Affymetrix platforms to control for some of the differences in microarray technology such as sensitivities, dynamic ranges and probe designs, etc. (Petersen et al. 2005, Tsai et al. 2006). We believe that the gene expression patterns that we observed between psychosis and non-psychosis cannot be explained by chance. One factor is the reasonable false discovery rate that would be expected at p-value<0.001. Another is the minimal overlap in genes found at each significance level between psychosis and bipolar disorder. For example, of the 1,281 genes in bipolar disorder (p<0.01), 6% are in common with psychosis, and of the 375 genes in bipolar disorder (p<0.001), 3% are in common with psychosis.

3. It seems important to discuss the relationship between these two analyses of the same data. They even got one of the same genes coming up (MT1X). It's rather confusing that these two analyses of the same data (in part) are giving the same result
when analyzed different ways, especially since the current study corrected for bipolar disease! (page 9) Surely this is something that requires an explanation.

We thank the Reviewer for pointing out this. MT1X is the only metallothionein gene found to be significant at the same low p-value threshold (p<0.001) in both bipolar disorder and psychosis. This is because MT1X gene is highly up-regulated in bipolar disorder with psychotic features, so that when adjusting for this gene with the regression model, the p-value is still less than 0.001 in psychosis. Nonetheless, other metallothionein genes found in bipolar disorder are at much more liberal p-value thresholds where the gene counts approach 3,300. Therefore the chance of an intersection of genes between psychosis and bipolar disorder is much higher.

4. Also pertaining to the shared subjects across datasets, Table 1 should be revised. As it currently reads, the reader can misinterpret ‘Subject Samples’ to mean the number of subjects under study, in which case the numbers would be incorrect. The row should be re-named ‘# of Samples’ and below it another row should be inserted to display the ‘# of Subjects’ (i.e. 82 and 81 for the Control and Psychosis groups respectively) for which the demographic data actually applies to. It would also be worth re-wording in the abstract ‘…over 400 individual microarray samples…’ to ‘…over 400 microarray samples comprised of 162 individual subjects…’.

We separated # of samples and # of subjects in Table 1 (page 30). We also changed the wording in the abstract as recommended (page 2).

5. The Methods section is a bit unclear. Did the authors use the pre-processed data, or did they use the raw data (.cel files) and run the same normalization methods across all datasets? In the Individual study analysis once the genes identified as significantly correlated with a demographic variable, it is not clear how this information was integrated with the psychosis analysis to obtain FC values. (p9. ‘….analysis of psychosis was performed…..adjusting for the demographic terms as listed above…)’? Please elaborate.

We have added additional explanation in the methods section. We have indicated "All microarray raw data (.cel files) were transformed using the Robust Multichip Average (RMA) normalization algorithm." (page 6). Regarding the psychosis analysis, we performed the individual study analysis adjusting for the significant demographic terms using a multiple regression model to obtain adjusted p-values and fold changes. We have rearranged the paragraph to clarify the individual study analysis procedure (page 8).

Minor essential revisions
6. In the discussion (pg 21) the authors state "Meta-analysis of schizophrenia shows that neuropeptide genes such as SST, NPY and TAC1 are not significantly regulated although these genes show a similar pattern of down-regulation as compared to the meta-analysis of psychosis". What meta-analysis of schizophrenia is being referred to? Furthermore, what does it mean for these genes to be not significantly regulated, but also showing down-regulation? This makes no sense to me.

We apologize for the lack of clarity. We found that neuropeptide genes such as SST, NPY and TAC1 are significantly down-regulated in the psychosis analysis. However, when we compared the results from the cross-study analysis of schizophrenia (SMRI online genomics dB) these neuropeptide genes were not statistically significant, although these genes showed a trend toward down-regulation in schizophrenia. We agree that this statement is rather confusing and doesn’t add much information to the manuscript so that we have removed these sentences in the revised manuscript.

7. Finally, the term meta-analysis seems somewhat inappropriate and even misleading for this study. The appeal of a meta-analysis is that it in effect combines findings from independent studies into one large study with many participants. In this study although there are multiple studies being combined, they are not independent, as they are all using samples from the same subject pool. If the authors use the term "combined analysis" or "cross-study analysis", there would be no problem.

We have changed “meta-analysis” to “cross-study analysis” or “combined analysis” throughout the manuscript.

Reviewer’s report 2

The authors describe in a clear and concise way their meta-analysis of gene expression microarray data, comparing subjects with and without psychotic symptoms. A dimensional rather than categorical organization of behavioural syndromes may correspond more closely to the neurobiology underlying psychotic symptoms, so their approach is warranted. These investigators are well-positioned in their role at the SMRI to conduct this study, and they find robust changes in gene expression associated with psychosis. The main findings are upregulated metallothionein genes in psychosis, which has been reported in only one previous paper, and decreased expression of neuropeptide RNAs, for which there have been earlier reports.

Overall, the results point to a novel set of candidate molecules that may be important in the neurobiology of psychosis. How this fits with the existing genetic and gene expression data implicating strong etiological candidate genes in schizophrenia remains unclear, and so further work is required to understand the mechanisms that connect metallothionein and neuropeptide function to that of genes like NRG1, DISC1 and
DOAO, as well as neurotransmitter systems involved in psychosis such as dopamine, glutamate and GABA. Overall, I feel this paper is an important contribution to the literature, and will be of interest to those in the field.

A few minor comments:

1. Page 20: “… NR4A2 knock-out mice suggested that these animals showed psychotic symptoms” To be precise, it would be better to state that these animals showed behaviours related to psychosis in humans or that behavioural endophenotypes resembled that seen in other animal models of psychosis/schizophrenia. Obviously one cannot assess core psychotic symptoms directly in any animal.

We thank the Reviewer for clarifying this statement. We have changed the statement as suggested (page 21).

2. Page 21: “Psychotic feature is an overlapping …” should be “Psychosis is an overlapping …”

We have removed the sentence because of the ambiguity of this statement in the discussion.

3. Page 15: “Most of individual studies show …” - delete “of”

We have made the change as suggested (page 14).

Reviewer’s report 3

This paper carries out a meta-analysis of a series of gene expression studies carried out on post-mortem brain samples from people with psychosis and controls. This is potentially very interesting. The authors claim that the meta-analysis increases the power of the studies to detect differentially expressed genes. This is likely to be the case but the analysis conducted here cannot be said to be a true meta-analysis and therefore these data are flawed.

1. The problem is that individual samples must have been included more than once in the meta-analysis and no account seems to have been taken of this. The two sources of tissue used for arraying are the Array Consortium (AC) and the Neuropathology consortium (NPC). The AC has 105 individual brains and the NPC 60. Table 2 indicates that studies 1, 3, 5 and 7 used BA46 samples from the AC, and the studies used 81, 86, 73 and 81 samples respectively. Therefore there must be overlap in the samples used
so effectively some of these samples are technical rather than biological replicates. In the analysis it appears that all samples have been given the same weight and there is no indication of the overlap between samples. Likewise the NPC has 55 individual brains and studies 2, 4 and 14 have used 55, 26 and 31 samples respectively.

We agree with the Reviewer that there are dependencies in samples between these studies, and that there requires adjustment in the analysis to make appropriate interpretations of the results. We implemented a correction factor to the degrees of freedom of each gene when calculating the p-values from the cross-study analysis, similar to our previous paper (Elashoff et al. 2007). This correction factor was identified using a permutation test on the 7 studies where we computed the exact degrees of freedom required to provide 5% of the gene significant at p<0.05. We have added a description of the permutation test on page 7.

2. Here different brain areas have been used so these are not direct technical replicates but they are not direct biological replicates either. The effect of the individual in, for instance, genetic variation, is likely to be reflected in all brain areas, but there will also be an effect of brain area on gene expression. Thus there are two levels of ambiguity in the analysis here and a direct meta-analysis is not appropriate.

We thank the Reviewer for pointing out this important issue. We are aware of the differences in gene expression in multiple brain regions such as different cortical regions as well as sub-cortical regions (Katsel et al. 2005, Haroutunian et al. 2007). However, to our knowledge, there is little information on the differences in gene expression in the sub-regions of the frontal cortex (BA46/10, BA8/9 and BA6) because most previous studies focused on one sub-region of the frontal cortex. One study showed that gene expression patterns between BA10 and BA46 are similar compared to non-frontal brain regions (Katsel et al. 2005).

In the previous paper (Elashoff et al. 2007), we included non-frontal region such as cerebellum for the combined study analysis. However, in the current study, we included the studies with only the frontal regions (5 studies with BA46, 1 study with BA8/9 and 1 study with BA6) for the reasons that we mentioned above. Please note that among the 7 frontal studies, the 5 studies with the prefrontal cortex (BA46) have larger subject numbers (N=81, 55, 86, 73 and 81) compared to the 2 studies with superior frontal cortex (N=26 and 31). Therefore, it is likely that the effects that we’ve seen in the psychosis analysis are mainly driven by the studies with BA46 region since there is a 2-3 fold difference in representation of this brain region.

Another issue was a limitation on tissue availability from the Neuropathology Consortium. This brain cohort has been widely used by many investigators for various types of research in the past. Therefore, there was a limitation with the prefrontal cortex (BA46) tissues when those microarray studies were conducted. However, these frontal regions including BA46/10, BA8/9 and BA6 are adjacent to each other and have been implicated in schizophrenia (Digney et al. 2005, Koch et al. 2007, Dean et al. 2007).
Please note that we also included another 4 studies that used the Array Collection with the same prefrontal cortex (BA46). We have added an explanation in the discussion section (page 17-18).

3. Given that the same biological samples are replicated several times this might increase the significance of real signals but will also increase the significance of any false positive signals. The QPCR has been carried out on the same samples so it is unlikely that it would reflect anything other than the results given from the arrays. If we believe the array results are accurate for each individual sample then indeed any other results would be unexpected. It is possible that I have misinterpreted the paper and the samples were independent but the information provided indicates that they are.

We agree with the Reviewer that we have used the same set of samples for the qPCR validation and it would have been more valuable if we had used an independent set of post-mortem brains. However, these two brain cohorts included in the current study are the best collections we have in terms of matched subjects. The Array Collection is suitable for this type of qPCR validation because of a large sample size (N=105). Although we included 4 microarray studies that used the tissues from the Array Collection, these studies did not have a complete set of subjects because of variable chip hybridization efficiency and post-chip quality control procedures. Therefore, we used a complete set of the Array Collection tissues (N=105) for the qPCR validation in order to minimize the issue with incomplete sample numbers in individual microarray studies. We believe that microarray technology offers a high-throughput and hypothesis-free approaches in candidate gene search in psychiatric research, but also more quantitative and direct method such as qPCR is required to confirm microarray findings due to inherent issues such as false discovery (Wang et al. 2006, Morey et al. 2006).

Please contact me with any additional questions concerning this manuscript. Thank you for providing us the opportunity to re-submit a revised manuscript and for your consideration.

Kwang Ho Choi, PhD
Stanley Laboratory of Brain Research
9800 Medical Center Dr. Bldg. 2C
Rockville, MD 20850 USA
References


